

Universal Quantization for Separate Encodings and Joint Decoding of Correlated Sources *

Avraham Reani and Neri Merhav
 Department of Electrical Engineering
 Technion - Israel Institute of Technology
 Technion City, Haifa 32000, Israel
 Emails: [avire@tx, merhav@ee].technion.ac.il

November 14, 2014

Abstract

We consider the multi-user lossy source-coding problem for continuous alphabet sources. In a previous work, Ziv proposed a single-user universal coding scheme which uses uniform quantization with dither, followed by a lossless source encoder (entropy coder). In this paper, we generalize Ziv's scheme to the multi-user setting. For this generalized universal scheme, upper bounds are derived on the redundancies, defined as the differences between the actual rates and the closest corresponding rates on the boundary of the rate region. It is shown that this scheme can achieve redundancies of no more than 0.754 bits per sample for each user. These bounds are obtained without knowledge of the multi-user rate region, which is an open problem in general. As a direct consequence of these results, inner and outer bounds on the rate-distortion achievable region are obtained.

Index Terms: Multi-terminal source coding, Dithered quantization, Universal source coding, scalar quantization, Slepian-Wolf coding.

*This paper was presented in part at 2014 IEEE International Symposium on Information Theory (ISIT).

*This research was supported by the Israeli Science Foundation (ISF), grant no. 208/08.

1 Introduction

Consider the case where two correlated sources are observed separately by two non-cooperative encoders which communicate with one decoder. The decoder needs to reconstruct both sources and the distortions between the reconstructions and the corresponding sources should not exceed some given values. The general version of this problem has remained open for several decades, even under the assumption of memoryless sources. However, many special cases have been solved. When no distortion is allowed, this is the problem considered by Slepian and Wolf [1]. Their well-known result states that two discrete sources X_1 and X_2 can be losslessly reproduced if and only if

$$R_1 \geq H(X_1|X_2), \tag{1a}$$

$$R_2 \geq H(X_2|X_1), \tag{1b}$$

$$R_1 + R_2 \geq H(X_1, X_2) \tag{1c}$$

where R_1 is the rate of the encoder observing X_1 and R_2 is the rate of the encoder observing X_2 . Returning to the lossy case, the setting in which one of the variables is known to the decoder, is the original Wyner-Ziv problem [2]. This setting was generalized to continuous alphabet sources by Wyner [3]. Other examples include the source coding problem with side information of Ahlswede-Körner [4], where an arbitrary distortion is allowed for one of the sources and the other source should be reconstructed losslessly. Berger and Yeung [5] considered a setting where one of the sources is to be perfectly reconstructed and the other source should be reconstructed with a distortion constraint (their setting subsumes all previous examples). Zamir and Berger [6] characterized the rate-distortion region in the high-SNR limit. Wagner and Anantharan [7] presented a new outer bound which is better than the previous outer bounds in the literature.

Recent results for specific sources and distortion measures include the works of Wagner, Tavildar, and Viswanath [8], who determined the rate region for the quadratic Gaussian multiterminal source coding problem, by showing that the Berger-Tung [9] inner bound is tight. In addition, a characterization of the rate region under logarithmic loss was given by Courtade and Weissman [10]. Finally, a version of this problem, where both users and the decoder must operate with zero-delay, was considered by Kaspi and Merhav [11], who characterized the rate region in this case.

In [12], Ziv presented a universal coding scheme for the single-user case. This scheme is composed of a uniform, one-dimensional quantizer with dither, followed by a noiseless variable-rate encoder (entropy encoder). He showed that this scheme yields a rate that is, for every positive integer n , no more than 0.754 bits per sample higher than the best possible rate associated with the optimal n -dimensional quantizer. This result was later revisited and further developed by Zamir and Feder [13], [14], who also gave a redundancy upper bound which depends on the source distribution. However, their derivation of the global upper bound relies on the known formula of the single-user rate-distortion function. In addition, a dithered scheme for the multi-user setting, which is similarly to the scheme in this paper, was given in [6]. Redundancy upper bounds can be derived by bounding the difference between the dithered scheme rate region and the outer bound on the multi-user rate region given in [6]. These bounds depend on the divergence between the source distribution and a Gaussian distribution. As a result, they are not uniformly bounded (for every source distribution) in contrast to the bound of Ziv and the bounds presented in this paper. In addition, only the redundancy of the sum of the rates can be upper bounded using the methods of [6].

In this paper, we investigate a generalized scheme for the multi-user setting. In this scheme, each user uses dithered quantizer followed by universal Slepian-Wolf encoder. We show that the rates achieved by this scheme are no more than 0.754 bits per sample away from the boundary of the achievable rate region, for each user. This is done regardless of the characterization of the achievable region, which is, as mentioned before, unknown in general. As a direct consequence of these results, inner and outer bounds on the achievable region are obtained. Finally, similarly to the results of [12], it is straightforward to show that using multi-dimensional lattice quantizers instead of scalar ones would decrease the redundancy to about 0.5 bits per sample for high lattice dimension.

The remainder of this paper is organized as follows. In Section 2, we present the problem formulation and give basic results regarding the performance of the dithered scheme. In Section 3, we revisit the redundancy upper bound of [12]. In Section 4, we enhance the results of Section 2 by adding an estimation stage to the dithered scheme. We conclude this work in Section 5.

2 Problem Formulation and Basic Results

Throughout the paper, random variables will be denoted by capital letters and their alphabets will be denoted by calligraphic letters. Random vectors (all of length n) will be denoted by capital letters in the bold face font.

In this section, we present the multi-user setting we deal with and describe the dithered coding scheme we use. Then, we give upper bounds on the performance of this scheme, compared to the boundary of the optimal rate region.

We begin with defining the multi-user rate region. Let (X_1, X_2) be a continuous alphabet memoryless source, characterized by the joint probability density $P_{X_1 X_2}$. We assume that $P_{X_1 X_2}$ has bounded support, i.e., there exists $A \in \mathbb{R}^+$ such that $P_{X_1 X_2}(x_1, x_2) = 0$ if $(x_1, x_2) \notin [-A, A] \times [-A, A]$. The reason for this assumption will be explained later. A rate pair (R_1^*, R_2^*) is said to be (D_1, D_2) -achievable under the mean-square error distortion measure with respect to (X_1, X_2) , if for every $\delta > 0$ and sufficiently large n , there exists a code of block length n consisting of two encoders f_1, f_2

$$f_1 : [-A, A]^n \rightarrow I_{M_1}, \quad f_2 : [-A, A]^n \rightarrow I_{M_2} \quad (2)$$

and a decoder g

$$g : I_{M_1} \times I_{M_2} \rightarrow [-A, A]^n \times [-A, A]^n \quad (3)$$

such that

$$\frac{1}{n} \mathbb{E} \|\mathbf{X}_1 - \hat{\mathbf{X}}_1\|^2 \leq D_1 + \delta, \quad \frac{1}{n} \mathbb{E} \|\mathbf{X}_2 - \hat{\mathbf{X}}_2\|^2 \leq D_2 + \delta \quad (4)$$

and

$$\frac{1}{n} \log M_1 \leq R_1^* + \delta, \quad \frac{1}{n} \log M_2 \leq R_2^* + \delta, \quad (5)$$

where $I_{M_i} \triangleq \{1, 2, \dots, M_i\}$, $i \in \{1, 2\}$. The set of (D_1, D_2) -achievable rate pairs, is denoted by $\mathcal{R}^*(D_1, D_2)$.

Our scheme works as follows. We have two encoders \tilde{f}_1, \tilde{f}_2 :

$$\tilde{f}_1 : [-A, A]^n \times [-\sqrt{3D_1}, \sqrt{3D_1}] \rightarrow I_{\tilde{M}_1}, \quad \tilde{f}_2 : [-A, A]^n \times [-\sqrt{3D_2}, \sqrt{3D_2}] \rightarrow I_{\tilde{M}_2} \quad (6)$$

and a decoder \tilde{g}

$$\tilde{g} : I_{\tilde{M}_1} \times I_{\tilde{M}_2} \times [-\sqrt{3D_1}, \sqrt{3D_1}] \times [-\sqrt{3D_2}, \sqrt{3D_2}] \rightarrow [-A, A]^n \times [-A, A]^n. \quad (7)$$

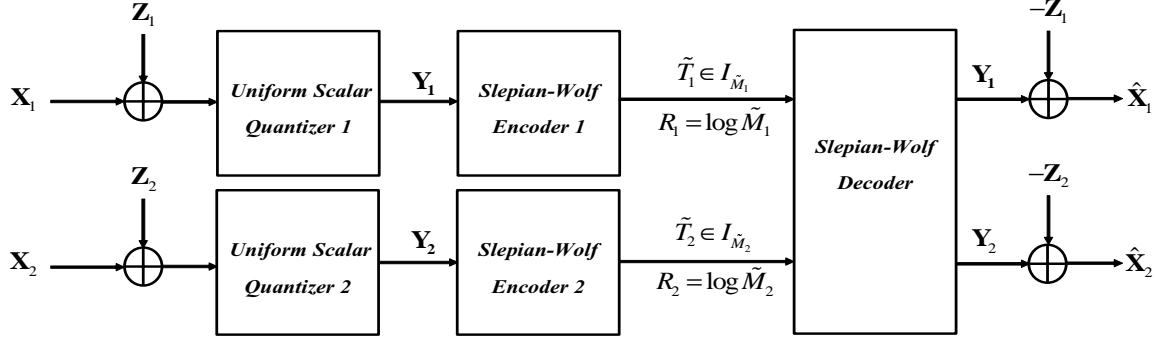


Figure 1: The dithered coding scheme

Each encoder \tilde{f}_i , $i \in \{1, 2\}$, uses a one-dimensional uniform quantizer \mathcal{Q}_i , $\mathcal{Q}_i : \mathbb{R} \rightarrow \{0, \pm 2\sqrt{3D_i}, \pm 2 \cdot 2\sqrt{3D_i}, \dots\}$ and a dither random variable (RV) Z_i , uniformly distributed over $[-\sqrt{3D_i}, \sqrt{3D_i}]$, to produce $\mathcal{Q}_i(\mathbf{X}_i + \mathbf{Z}_i) \triangleq [\mathcal{Q}_i(X_{i,1} + Z_i), \mathcal{Q}_i(X_{i,2} + Z_i), \dots, \mathcal{Q}_i(X_{i,n} + Z_i)]$, where \mathbf{Z}_i denotes a vector of dimension n composed of n repetitions of the same realization of Z_i . For convenience, the random variable $\mathcal{Q}_i(X_i + Z_i)$ and the random vector $\mathcal{Q}_i(\mathbf{X}_i + \mathbf{Z}_i)$ will be denoted by Y_i and \mathbf{Y}_i , respectively. The dither RV's, Z_1 and Z_2 , are available to the respective encoders and to the decoder and are independent. As is shown in [12, Lemma 1],

$$\mathbb{E} [||Y_i - Z_i - X_i||^2 | X_i] = D_i, \quad i \in \{1, 2\} \quad (8)$$

where the expectation is taken over Z_i . Notice that the distortion is D_i independently of X_i and therefore the total distortion is also D_i . After the quantization stage, the two encoders perform Slepian-Wolf encoding with a rate pair (R_1, R_2) , for lossless compression of \mathbf{Y}_1 and \mathbf{Y}_2 . Complying with Eq. (1), the rate pair $(R_1 = \log \tilde{M}_1, R_2 = \log \tilde{M}_2)$ satisfies

$$R_1 \geq H(Y_1 | Y_2, Z_1, Z_2), \quad (9a)$$

$$R_2 \geq H(Y_2 | Y_1, Z_1, Z_2), \quad (9b)$$

$$R_1 + R_2 \geq H(Y_1, Y_2 | Z_1, Z_2) \quad (9c)$$

where we used the following, for every value of n

$$\frac{1}{n} H(\mathbf{Y}_1 | \mathbf{Y}_2, Z_1, Z_2) = H(Y_1 | Y_2, Z_1, Z_2), \quad (10a)$$

$$\frac{1}{n} H(\mathbf{Y}_2 | \mathbf{Y}_1, Z_1, Z_2) = H(Y_2 | Y_1, Z_1, Z_2), \quad (10b)$$

$$\frac{1}{n} H(\mathbf{Y}_1, \mathbf{Y}_2 | Z_1, Z_2) = H(Y_1, Y_2 | Z_1, Z_2). \quad (10c)$$

To see why (10a) is true, consider the following chain

$$\begin{aligned}
\frac{1}{n}H(\mathbf{Y}_1|\mathbf{Y}_2, Z_1, Z_2) &= \frac{1}{n} \sum_{i=1}^n H(Y_{1,i}|Y_{1,1}, Y_{1,2}, \dots, Y_{1,i-1}, \mathbf{Y}_2, Z_1, Z_2) \\
&= \frac{1}{n} \sum_{i=1}^n H(Y_{1,i}|Y_{2,i}, Z_1, Z_2) \\
&= H(Y_1|Y_2, Z_1, Z_2)
\end{aligned} \tag{11}$$

where the second equality stems from the fact that Y_1 and Y_2 are memoryless given Z_1 and Z_2 and the third equality stems from the stationarity of the source. The same can be done for $H(\mathbf{Y}_2|\mathbf{Y}_1, Z_1, Z_2)$ and $H(\mathbf{Y}_2, \mathbf{Y}_1|Z_1, Z_2)$.

The rate region of Eq. (9) is achievable for n sufficiently large and it is denoted by $\mathcal{R}(D_1, D_2)$. The interesting range of R_1 is $\overline{\mathcal{R}}_1(D_1, D_2) \triangleq [H(Y_1|Y_2, Z_1, Z_2), H(Y_1|Z_1)]$ since higher rate can always be reduced to this range. The same is true for R_2 . The universal decoder first decodes \mathbf{Y}_1 and \mathbf{Y}_2 (correctly with high probability), and then subtracts the corresponding dithers to obtain the reconstruction vectors $\hat{\mathbf{X}}_1, \hat{\mathbf{X}}_2$:

$$\hat{\mathbf{X}}_i = \mathbf{Y}_i - \mathbf{Z}_i. \tag{12}$$

The universal Slepian-Wolf decoder is described in Appendix A. The dithered coding scheme is presented in Fig. 1.

Remark. The Slepian-Wolf mechanism can be applied, in general, to sources with countably-infinite alphabets. However, a universal Slepian-Wolf scheme for such sources is not known. Trying to preserve universality in the case of infinite alphabets would require the assignment of infinite number of sequences into bins. Thus, even the codebook generation does not seem to be feasible in this case. This is not surprising, considering the fact that even in the single-user case, diminishing redundancy cannot be achieved for universal lossless coding of sources with infinite alphabets (see, e.g., [15]). Therefore, for the sake of universality, we assumed that the source alphabets have bounded supports so the outputs of the quantizers have finite alphabets. From the above, this assumption is also needed for the original single-user scheme of Ziv [12]. The inner and outer bounds on the achievable rate-distortion region, which are obtained as a direct consequence of Theorems 1-4 below, are also valid, of course, for sources with unbounded support, as they do not depend on the universality.

We begin with a simple result.

Theorem 1. *For any rate pair (R_1^*, R_2^*) on the boundary of $\mathcal{R}^*(D_1, D_2)$ and any rate pair*

(R_1, R_2) on the boundary of $\mathcal{R}(D_1, D_2)$, with $R_1 \in \overline{\mathcal{R}}_1(D_1, D_2)$, we have

$$R_1 + R_2 \leq R_1^* + R_2^* + 2c \quad (13)$$

where $c = 0.754$ bits/sample.

Moreover, for any $R_1^* \in \overline{\mathcal{R}}_1(D_1, D_2)$, there exists a rate pair $(R_1, R_2) \in \mathcal{R}(D_1, D_2)$ such that

$$\begin{aligned} R_1 &= R_1^* \\ R_2 &\leq R_2^* + 2c. \end{aligned} \quad (14)$$

Proof of Theorem 1. We have

$$\begin{aligned} & \frac{1}{n} H(\mathbf{Y}_1, \mathbf{Y}_2 | Z_1, Z_2) \\ & \leq \frac{1}{n} H(\mathbf{Y}_1, \mathbf{Y}_2, T_1, T_2 | Z_1, Z_2) \\ & \leq \frac{1}{n} H(T_1, T_2) + \frac{1}{n} H(\mathbf{Y}_1, \mathbf{Y}_2 | T_1, T_2, Z_1, Z_2) \\ & \leq R_1^* + R_2^* + \frac{1}{n} H(\mathbf{Y}_1, \mathbf{Y}_2 | T_1, T_2, Z_1, Z_2) \\ & \leq R_1^* + R_2^* + \frac{1}{n} H(\mathbf{Y}_1, \mathbf{Y}_2 | g(T_1, T_2), Z_1, Z_2) \\ & = R_1^* + R_2^* + \frac{1}{n} H(\mathbf{Y}_1, \mathbf{Y}_2 | \hat{\mathbf{X}}_1^{opt}, \hat{\mathbf{X}}_2^{opt}, Z_1, Z_2) \\ & \leq R_1^* + R_2^* + \frac{1}{n} H(\mathbf{Y}_1 | \hat{\mathbf{X}}_1^{opt}, Z_1) \\ & \quad + \frac{1}{n} H(\mathbf{Y}_2 | \hat{\mathbf{X}}_2^{opt}, Z_2) \\ & \leq R_1^* + R_2^* + 2c \end{aligned} \quad (15)$$

where $T_1 \in I_{M_1}$, $T_2 \in I_{M_2}$ are the outputs of the optimal encoders f_1 , f_2 , respectively, $(\hat{\mathbf{X}}_1^{opt}, \hat{\mathbf{X}}_2^{opt}) \triangleq g(T_1, T_2)$ are the outputs of the optimal decoder g , and $(R_1^*, R_2^*) \in \mathcal{R}^*(D_1, D_2)$. The last inequality can be obtained in the same way as in [12]. The left-hand side is achievable for sufficiently large n . Therefore, for any rate pair $(R_1, R_2) \in \mathcal{R}(D_1, D_2)$, which lies on the straight line $R_1 + R_2 = H(Y_1, Y_2 | Z_1, Z_2)$, we have

$$R_1 + R_2 \leq R_1^* + R_2^* + 2c \quad (16)$$

Moreover, if $R_1^* \in \overline{\mathcal{R}}_1(D_1, D_2)$, we can always take $R_1 = R_1^*$ and obtain:

$$R_2 \leq R_2^* + 2c \quad (17)$$

The same can be done, of course, when the roles of the two users are interchanged. This completes the proof. \square

The following theorem suggests another result regarding the relation between the boundary of $\mathcal{R}(D_1, D_2)$ and that of $\mathcal{R}^*(D_1, D_2)$.

Theorem 2. *For any rate pair (R_1, R_2) on the boundary of $\mathcal{R}(D_1, D_2)$, with $R_1 \in \overline{\mathcal{R}}_1(D_1, D_2)$, there exists a rate pair $(R_1^*, R_2^*) \in \mathcal{R}^*(D_1, D_2)$ such that:*

$$\begin{aligned} R_1 &\leq R_1^* + c \\ R_2 &\leq R_2^* + c \end{aligned} \tag{18}$$

Notice that Theorems 1 and 2 also provide outer bounds on $\mathcal{R}^*(D_1, D_2)$. Theorem 1 asserts that the straight line $R_1 + R_2 = H(Y_1, Y_2 | Z_1, Z_2) - 2c$ defines an outer bound for $\mathcal{R}^*(D_1, D_2)$. In addition, Theorem 2 bounds the distance between the boundary of $\mathcal{R}(D_1, D_2)$ and that of $\mathcal{R}^*(D_1, D_2)$ in each coordinate. The boundary of $\mathcal{R}(D_1, D_2)$ is, of course, an inner bound on $\mathcal{R}^*(D_1, D_2)$.

Before proving Theorem 2, we first prove a simple auxiliary result regarding the source-coding problem where side information is available only to the encoders but not to the decoder. The setting is as follows. A rate pair (R_1, R_2) is achievable for a memoryless source $(\mathcal{Y}_1, \mathcal{Y}_2, P_{Y_1, Y_2})$ and some side information $S \in \mathcal{S}$ which depends statistically on $(\mathbf{Y}_1, \mathbf{Y}_2)$ through the joint probability distributions $P_{\mathbf{Y}_1, \mathbf{Y}_2, S}$, if for any $\delta > 0$ and sufficiently large n , there exists a block code of length n consisting of two encoders f_1, f_2

$$f_1 : \mathcal{Y}_1^n \times \mathcal{S} \rightarrow I_{M_1}, \quad f_2 : \mathcal{Y}_2^n \times \mathcal{S} \rightarrow I_{M_2} \tag{19}$$

and a decoder g

$$g : I_{M_1} \times I_{M_2} \rightarrow \mathcal{Y}_1^n \times \mathcal{Y}_2^n \tag{20}$$

such that

$$\Pr\{g(f_1(\mathbf{Y}_1, S), f_2(\mathbf{Y}_2, S)) \neq (\mathbf{Y}_1, \mathbf{Y}_2)\} \leq \delta \tag{21}$$

and

$$\frac{1}{n} \log M_1 \leq R_1 + \delta, \quad \frac{1}{n} \log M_2 \leq R_2 + \delta \tag{22}$$

The set of achievable rate pairs is denoted by $\tilde{\mathcal{R}}$. The regular Slepian-Wolf region (without side information) is denoted by \mathcal{R}_{SW} . Obviously, $\mathcal{R}_{SW} \subseteq \tilde{\mathcal{R}}$. We have the following lemma.

Lemma 1. *Any rate pair $(\tilde{R}_1, \tilde{R}_2) \in \tilde{\mathcal{R}}$ must satisfy the following constraint:*

$$\tilde{R}_1 + \tilde{R}_2 \geq H(Y_1, Y_2). \quad (23)$$

Therefore, side information available only to the encoders cannot improve the performance if $\tilde{R}_1 \in [H(Y_1|Y_2), H(Y_1)]$ or $\tilde{R}_2 \in [H(Y_2|Y_1), H(Y_2)]$.

Proof of Lemma 1. The proof follows directly from the fact that even one encoder, which has access to $(\mathbf{Y}_1, \mathbf{Y}_2, S)$, cannot do better than $H(\mathbf{Y}_1, \mathbf{Y}_2)$, when the side information S is not available to the decoder. \square

The generalization of Lemma 1 to our case where, in addition, a dither is available to the encoders and decoder, is straightforward. We can now prove Theorem 2.

Proof of Theorem 2. Assume that the optimal code (f_1, f_2, g) , which achieves the rate pair (R_1^*, R_2^*) , is known, and that the encoders of the dithered scheme, which transmit $\mathbf{Y}_1, \mathbf{Y}_2$ at rates (R_1, R_2) to the decoder, have access to $f_1(\mathbf{X}_1), f_2(\mathbf{X}_2)$ as side information. According to Lemma 1, this side information does not change the fact that any rate pair $(R_1, R_2) \in \mathcal{R}(D_1, D_2)$ must satisfy $R_1 + R_2 \geq H(Y_1, Y_2|Z_1, Z_2)$. Consider the following auxiliary coding scheme: User i compresses $T_i = f_i(\mathbf{X}_i)$ using nR_i^* bits, $i \in \{1, 2\}$. Then, the first user uses Slepian-Wolf coding to compress \mathbf{Y}_1 given $\{T_1, T_2, Z_1\}$ into $H(\mathbf{Y}_1|T_1, T_2, Z_1)$ bits. The second user uses Slepian-Wolf coding to compress \mathbf{Y}_2 given $\{\mathbf{Y}_1, T_1, T_2, Z_1, Z_2\}$ into $H(\mathbf{Y}_2|\mathbf{Y}_1, T_1, T_2, Z_1, Z_2)$ bits. The decoder, which has access to $\{T_1, T_2, Z_1, Z_2\}$, first decodes \mathbf{Y}_1 , using $\{T_1, T_2, Z_1\}$. Then, it decodes \mathbf{Y}_2 using $\{\mathbf{Y}_1, T_1, T_2, Z_1, Z_2\}$. The rate pair of this scheme, (R_1, R_2) , satisfies

$$\begin{aligned} R_1 &= R_1^* + \frac{1}{n} H(\mathbf{Y}_1|T_1, T_2, Z_1) \\ &\leq R_1^* + \frac{1}{n} H(\mathbf{Y}_1|g(T_1, T_2), Z_1) \\ &= R_1^* + \frac{1}{n} H(\mathbf{Y}_1|\hat{\mathbf{X}}_1^{opt}, \hat{\mathbf{X}}_2^{opt}, Z_1) \\ &\leq R_1^* + \frac{1}{n} H(\mathbf{Y}_1|\hat{\mathbf{X}}_1^{opt}, Z_1) \end{aligned} \quad (24)$$

and

$$\begin{aligned}
R_2 &= R_2^* + \frac{1}{n} H(\mathbf{Y}_2 | \mathbf{Y}_1, T_1, T_2, Z_1, Z_2) \\
&\leq R_2^* + \frac{1}{n} H(\mathbf{Y}_2 | \mathbf{Y}_1, g(T_1, T_2), Z_1, Z_2) \\
&= R_2^* + \frac{1}{n} H(\mathbf{Y}_2 | \mathbf{Y}_1, \hat{\mathbf{X}}_1^{opt}, \hat{\mathbf{X}}_2^{opt}, Z_1, Z_2) \\
&\leq R_2^* + \frac{1}{n} H(\mathbf{Y}_2 | \hat{\mathbf{X}}_2^{opt}, Z_2)
\end{aligned} \tag{25}$$

The upper bounds on $H(\mathbf{Y}_i | \hat{\mathbf{X}}_i^{opt}, Z_i)$ can be obtained in the same way as in [12]. Notice that the Slepian-Wolf coding part in the proof requires long blocks of $(T_1, T_2, \mathbf{Y}_1, \mathbf{Y}_2)$. Now, since $\mathcal{R}(D_1, D_2) \subseteq \mathcal{R}^*(D_1, D_2)$, we can always find R (R_1^*, R_2^*) $\in \mathcal{R}^*(D_1, D_2)$ such that $R_1^* + c \in \overline{\mathcal{R}}_1(D_1, D_2)$ (or higher and thus can be reduced to this range). Using the auxiliary scheme above, the rate pair $(R_1, R_2) = (R_1^* + c, R_2^* + c)$ can be achieved. Therefore, it can also be achieved by the dithered scheme, since $R_1 \in \overline{\mathcal{R}}_1(D_1, D_2)$ (or higher), and in this range the regions of the auxiliary scheme and the dithered scheme coincide. Notice that any rate pair in $\mathcal{R}(D_1, D_2)$ can be achieved in practice by time-sharing the two edge points of $\mathcal{R}(D_1, D_2)$. \square

3 Revisiting the Upper Bound on $H(\mathbf{Y} | \hat{\mathbf{X}}^{opt}, \mathbf{Z})$

In this section, we revisit the proof of [12] for the upper bound on $H(\mathbf{Y} | \hat{\mathbf{X}}^{opt}, \mathbf{Z})$. This is done for completeness and since we point and modify some of the steps in the next section. The result of this section involves only one source X . The width of the quantization cell is denoted by $\Delta \triangleq 2\sqrt{3D} \Rightarrow D = \Delta^2/12$.

First, we show that for each coordinate X_k , $k \in \{1, \dots, n\}$,

$$\mathbb{E}[X_k - \hat{X}_k^{opt} + Z] = 0. \tag{26}$$

This follows from the following consideration:

$$\begin{aligned}
\mathbb{E}[X_k - \hat{X}_k^{opt} + Z] &= \mathbb{E}[X_k - \hat{X}_k^{opt}] + \mathbb{E}[Z] \\
&= \mathbb{E}[X_k - \hat{X}_k^{opt}].
\end{aligned} \tag{27}$$

The distortion associated with X_k is given by:

$$\begin{aligned}
\mathbb{E}\left[\left(X_k - \hat{X}_k^{opt}\right)^2\right] &= \text{Var}\{X_k - \hat{X}_k^{opt}\} + \left(\mathbb{E}[X_k - \hat{X}_k^{opt}]\right)^2 \\
&\geq \text{Var}\{X_k - \hat{X}_k^{opt}\}
\end{aligned} \tag{28}$$

where the inequality must be achieved by the optimal quantizer. Otherwise, we could add a constant to \hat{X}_k^{opt} to obtain $\mathbb{E}[X_k - \hat{X}_k^{opt}] = 0$ and thus smaller total distortion, in contradiction to the optimality of the quantizer.

We now rederive the upper bound on $H(\mathbf{Y}|\mathbf{X}^{opt}, \mathbf{Z})$. Using a method similar to [13], we show the following for the conditional entropy of each coordinate:

$$\begin{aligned} H(Y_k|\hat{X}_k^{opt}, Z) &= I(X_k; X_k + Z|\hat{X}_k^{opt}) \\ &= h(X_k + Z|\hat{X}_k^{opt}) - h(Z) \end{aligned} \quad (29)$$

where the second equality follows since \hat{X}_k^{opt} and Z are independent. By definition:

$$\begin{aligned} H(Y_k|\hat{X}_k^{opt} = q, Z) &= \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} dz f_Z(z) H(Y_k|\hat{X}_k^{opt} = q, Z = z) \\ &= \frac{1}{\Delta} \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} dz H(Y_k|\hat{X}_k^{opt} = q, Z = z) \end{aligned} \quad (30)$$

Given $(\hat{X}_k^{opt} = q, Z = z)$, Y_k is a discrete random variable taking values in $\{i\Delta\}_{i \in \mathbb{N}}$. Thus,

$$\begin{aligned} H(Y_k|\hat{X}_k^{opt} = q, Z = z) &= - \sum_{j \in \mathbb{N}} P_{Y_k|\hat{X}_k^{opt}, Z}(j\Delta|q, z) \\ &\quad \cdot \log P_{Y_k|\hat{X}_k^{opt}, Z}(j\Delta|q, z) \end{aligned} \quad (31)$$

where $P_{Y_k|\hat{X}_k^{opt}, Z}(\cdot|q, z)$ is the probability density function of Y_k given \hat{X}_k^{opt} and Z . Calculating:

$$\begin{aligned} P_{Y_k|\hat{X}_k^{opt}, Z}(j\Delta|q, z) &= \Pr\{Y_k = j\Delta|Z = z, \hat{X}_k^{opt} = q\} \\ &= \Delta \int_{(j-\frac{1}{2})\Delta-z}^{(j+\frac{1}{2})\Delta-z} dx \frac{1}{\Delta} f_{X|\hat{X}_k^{opt}}(x|q) \\ &= \Delta \cdot f_{U_k|\hat{X}_k^{opt}}(j\Delta - z|q) \end{aligned} \quad (32)$$

where $f_{X|\hat{X}_k^{opt}}(\cdot|q)$ is the probability density function of X given \hat{X}_k^{opt} , $f_{U_k|\hat{X}_k^{opt}}(\cdot|q) = f_{X|\hat{X}_k^{opt}}(\cdot|q) * f_Z(\cdot)$ is the probability density function of the continuous random variable $U_k \triangleq X_k + Z$ given \hat{X}_k^{opt} and $'*'$ denotes the convolution operation. Substituting in Eq.

(30), we have

$$\begin{aligned}
H(Y_k | \hat{X}_k^{opt} = q, Z) &= -\frac{1}{\Delta} \int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} dz \sum_{i \in \mathbb{N}} \Delta \cdot f_{U_k | \hat{X}_k^{opt}}(j\Delta - z|q) \\
&\quad \cdot \log \left(\Delta \cdot f_{U_k | \hat{X}_k^{opt}}(j\Delta - z|q) \right) \\
&= -\int_{\mathbb{R}} du \cdot f_{U_k | \hat{X}_k^{opt}}(u|q) \cdot \log \left(\Delta \cdot f_{U_k | \hat{X}_k^{opt}}(u|q) \right) \\
&= h(U_k | \hat{X}_k^{opt} = q) - \log \Delta \\
&= h(U_k | \hat{X}_k^{opt} = q) - h(Z) \\
&= h(U_k | \hat{X}_k^{opt} = q) - h(Z | X_k, \hat{X}_k^{opt} = q) \\
&= h(U_k | \hat{X}_k^{opt} = q) - h(U_k | X_k, \hat{X}_k^{opt} = q) \\
&= I(X_k; X_k + Z | \hat{X}_k^{opt} = q)
\end{aligned} \tag{33}$$

where in the fifth equality we used the independence of X_k and Z and in the sixth equality we used the fact that $U_k = X_k + Z$. We have

$$\begin{aligned}
H(Y_k | \hat{X}_k^{opt}, Z) &= \sum_{q \in \mathcal{Q}_{opt}} P_{\hat{X}_k^{opt}}(q) H(Y_k | \hat{X}_k^{opt} = q, Z) \\
&= \sum_{q \in \mathcal{Q}_{opt}} P_{\hat{X}_k^{opt}}(q) I(X_k; X_k + Z | \hat{X}_k^{opt} = q) \\
&= I(X_k; X_k + Z | \hat{X}_k^{opt}) \\
&= h(X_k + Z | \hat{X}_k^{opt}) - h(Z) \\
&= h(X_k - \hat{X}_k^{opt} + Z | \hat{X}_k^{opt}) - h(Z)
\end{aligned} \tag{34}$$

This completes the derivation of Eq. (29). Now, we can upper bound $h(X_k - \hat{X}_k^{opt} + Z | \hat{X}_k^{opt})$ in the following way.

$$\begin{aligned}
h(X_k - \hat{X}_k^{opt} + Z | \hat{X}_k^{opt}) &= \sum_{q \in \mathcal{Q}_{opt}} P_{\hat{X}_k^{opt}}(q) h(X - \hat{X}_k^{opt} + Z | \hat{X}_k^{opt} = q) \\
&\leq \sum_{q \in \mathcal{Q}_{opt}} P_{\hat{X}_k^{opt}}(q) \\
&\quad \cdot \frac{1}{2} \log \left(2\pi e \mathbb{E} \left[(X - \hat{X}_k^{opt} + Z)^2 | \hat{X}_k^{opt} = q \right] \right) \\
&\leq \frac{1}{2} \log \left(2\pi e \mathbb{E} (X - \hat{X}_k^{opt} + Z)^2 \right)
\end{aligned} \tag{35}$$

where in the first inequality we upper bounded the differential entropy by using the maximum-entropy property of the Gaussian random variable and the second inequality is due to Jensen.

Using these results, we can upper bound $H(\mathbf{Y}|\hat{\mathbf{X}}^{opt}, \mathbf{Z})$.

$$\begin{aligned}
H(\mathbf{Y}|\hat{\mathbf{X}}^{opt}, \mathbf{Z}) &\leq \sum_{k=1}^n H(Y_k|\hat{X}_k^{opt}, Z) \\
&\leq \sum_{k=1}^n \frac{1}{2} \log \left(2\pi e \mathbb{E} \left[(X_k - \hat{X}_k^{opt} + Z)^2 \right] \right) \\
&\quad -nh(Z) \\
&\leq \frac{n}{2} \log \left(2\pi e \frac{1}{n} \sum_{k=1}^n \mathbb{E} \left[(X_k - \hat{X}_k^{opt} + Z)^2 \right] \right) \\
&\quad -n \log \Delta \\
&= \frac{n}{2} \log \left(2\pi e \frac{1}{n} \mathbb{E} \left\| \mathbf{X} - \hat{\mathbf{X}}^{opt} + \mathbf{Z} \right\|^2 \right) \\
&\quad -n \log \Delta \\
&\leq \frac{n}{2} \log (2\pi e 2D) - n \log \Delta \\
&= \frac{n}{2} \log (2\pi e 2D) - \frac{n}{2} \log (\Delta^2) \\
&= \frac{n}{2} \log (4\pi e D) - \frac{n}{2} \log (12D) \\
&= \frac{n}{2} \log \left(\frac{\pi e}{3} \right)
\end{aligned} \tag{36}$$

where the third inequality is due to Jensen, and in the fourth we used the following.

$$\begin{aligned}
\frac{1}{n} \mathbb{E} \left\| \mathbf{X} - \hat{\mathbf{X}}^{opt} + \mathbf{Z} \right\|^2 &= \frac{1}{n} \mathbb{E} \left\| \mathbf{X} - \hat{\mathbf{X}}^{opt} \right\|^2 + \frac{1}{n} \mathbb{E} \left\| \mathbf{Z} \right\|^2 \\
&\leq 2D
\end{aligned} \tag{37}$$

which stems from the independence of X and Z . This completes the proof of the upper bound on $H(\mathbf{Y}|\hat{\mathbf{X}}^{opt}, \mathbf{Z})$.

4 Improving the Bounds by Adding an Estimation Stage

The goal of this section is to enhance the results of Section 1 by improving the coding scheme described there. The idea is to decrease the distortion by adding an estimation stage at the decoder side. The new scheme works as follows. After producing $\mathbf{Y}_1, \mathbf{Y}_2$ and instead of just using them as outputs, the decoder uses them to estimate each one of the source vectors $(\mathbf{X}_1, \mathbf{X}_2)$. Since the sources and the quantization process (given Z) are memoryless, the estimation can be done on a symbol-by-symbol basis.

We begin with the following lemma:

Lemma 2. *For the multi-terminal setting described in Section 1, we have ($i \in \{1, 2\}$):*

$$\mathbb{E}[Y_i - Z_i] = \mathbb{E}[X_i] \quad (38)$$

$$\mathbb{E}[(Y_i - Z_i)^2] = \mathbb{E}[X_i^2] + D_i \quad (39)$$

$$\mathbb{E}[X_i(Y_i - Z_i)] = \mathbb{E}[X_i^2] \quad (40)$$

$$\mathbb{E}[(Y_1 - Z_1)(Y_2 - Z_2)] = \mathbb{E}[X_1 X_2] \quad (41)$$

$$\mathbb{E}[X_1(Y_2 - Z_2)] = \mathbb{E}[X_1 X_2] \quad (42)$$

$$\mathbb{E}[X_2(Y_1 - Z_1)] = \mathbb{E}[X_1 X_2] \quad (43)$$

Notice that the results above are true for each coordinate $k \in \{1, \dots, n\}$. The proof of Lemma 2 is given in Appendix B.

The improved decoder described below requires the knowledge of the second-order statistics of the source. However, as Lemma 2 shows, these statistics can be estimated from $\{\mathbf{Y}_i\}_{i=1}^2$, so universality can still be maintained.

The decoder of the multi-terminal setting uses the optimal linear estimator, under the MMSE criterion, of $\{X_i\}_{i=1}^2$ given $\{Q(X_i + Z_i) - Z_i\}_{i=1}^2$. The estimation error is calculated by using the results of Lemma 2. From now on, without loss of generality, we assume that $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 0$. The covariance matrix of $\underline{Y} \triangleq [Q(X_1 + Z_1) - Z_1, Q(X_2 + Z_2) - Z_2]$ is:

$$\Lambda = \begin{pmatrix} \mathbb{E}[X_1^2] + D_1 & \mathbb{E}[X_1 X_2] \\ \mathbb{E}[X_1 X_2] & \mathbb{E}[X_2^2] + D_2 \end{pmatrix} \quad (44)$$

and the inverse matrix is:

$$\Lambda^{-1} = \frac{1}{|\Lambda|} \begin{pmatrix} \mathbb{E}[X_2^2] + D_2 & -\mathbb{E}[X_1 X_2] \\ -\mathbb{E}[X_1 X_2] & \mathbb{E}[X_1^2] + D_1 \end{pmatrix} \quad (45)$$

The vector $\mathbb{E}[X_1 \cdot \underline{Y}^\dagger]$ is given by:

$$\mathbb{E}[X_1 \cdot \underline{Y}^\dagger] = \begin{pmatrix} \mathbb{E}[X_1^2] \\ \mathbb{E}[X_1 X_2] \end{pmatrix} \quad (46)$$

It can be shown by direct calculation that

$$\Lambda^{-1} \mathbb{E}[X_1 \cdot \underline{Y}^\dagger] = \frac{1}{|\Lambda|} \begin{pmatrix} |\Lambda| - D_1(\mathbb{E}[X_2^2] + D_2) \\ \mathbb{E}[X_1 X_2] D_1 \end{pmatrix}$$

Therefore, the optimal linear estimator of X_1 given the vector \underline{Y} is:

$$\hat{X}_1 = \underline{Y} \cdot \frac{1}{|\Lambda|} \begin{pmatrix} |\Lambda| - D_1(\mathbb{E}[X_2^2] + D_2) \\ \mathbb{E}[X_1 X_2] D_1 \end{pmatrix} \quad (47)$$

The error of the optimal linear estimator is given by:

$$D_1^* = \mathbb{E}[X_1^2] - \mathbb{E}[\hat{X}_1^2] \quad (48)$$

It is shown in Appendix C that the estimation error takes the following form:

$$D_1^* = D_1 \frac{\mathbb{E}[X_1^2](\mathbb{E}[X_2^2] + D_2) - \mathbb{E}[X_1 X_2]^2}{(\mathbb{E}[X_1^2] + D_1)(\mathbb{E}[X_2^2] + D_2) - \mathbb{E}[X_1 X_2]^2}$$

Remember that D_1^* is the distortion of X_1 in the multi-terminal setting, where we add the above estimation stage after decoding $(\mathbf{Y}_1, \mathbf{Y}_2)$. It can be easily seen that the fraction in the brackets is less than 1 and thus $D_1^* \leq D_1$ as desired. The same can be done, of course, for X_2 . Since the distortion of X_i in the improved scheme is D_i^* , we should compare the rate pair (R_1, R_2) of this scheme, to the optimal rate pair (R_1^*, R_2^*) which achieves (D_1^*, D_2^*) . This fact immediately improves on the results of Theorems 1 and 2. Revisiting the derivation of the upper bound for $H(\mathbf{Y}|\hat{\mathbf{X}}^{opt}, \mathbf{Z})$ in Eq. (36), it can be shown that ($i \in \{1, 2\}$):

$$H(\mathbf{Y}_i|\hat{\mathbf{X}}_i^{opt}, \mathbf{Z}_i) \leq \frac{n}{2} \log \left[\frac{\pi e}{6} \left(\frac{D_i^*}{D_i} + 1 \right) \right] \quad (49)$$

by using the following:

$$\begin{aligned} \frac{1}{n} \mathbb{E} \left\| \mathbf{X}_i - \hat{\mathbf{X}}_i^{opt} + \mathbf{Z}_i \right\|^2 &= \frac{1}{n} \mathbb{E} \left\| \mathbf{X}_i - \hat{\mathbf{X}}_i^{opt} \right\|^2 \\ &+ \frac{1}{n} \mathbb{E} \left\| \mathbf{Z}_i \right\|^2 \\ &\leq D_i^* + D_i \end{aligned} \quad (50)$$

Notice that when X_1 and X_2 are independent, $\mathbb{E}[X_1 X_2] = 0$ and we have

$$H(\mathbf{Y}_i|\hat{\mathbf{X}}_i^{opt}, \mathbf{Z}_i) \leq \frac{n}{2} \log \left[\frac{\pi e}{6} \left(2 - \frac{D_i}{\mathbb{E}[X_i^2] + D_i} \right) \right] \quad (51)$$

The maximum interesting value of D_i^* is, of course, $\mathbb{E}[X_i^2]$. This value is obtained for $D_i \rightarrow \infty$. It is not hard to see that the range of the upper bound in (51) is $[0.255, 0.755]$ and that it is a decreasing function of D_1 . For the high-SNR limit, i.e., $D_i \rightarrow 0$, it is well known that the redundancy is 0.255 bits/sample (cf. [16]). We define ($i \in \{1, 2\}$):

$$c_i(D_1, D_2) = \frac{n}{2} \log \left[\frac{\pi e}{6} \left(\frac{D_i^*}{D_i} + 1 \right) \right] \quad (52)$$

We can now state Theorems 3 and 4. These theorems are obtained by applying the generalized upper bound of Eq. (49), instead of Ziv's upper bound on $H(\mathbf{Y}_i|\hat{\mathbf{X}}_i^{opt}, \mathbf{Z}_i)$, in the proofs of Theorem 1 and 2.

Theorem 3. For any rate pair (R_1^*, R_2^*) on the boundary of $\mathcal{R}^*(D_1, D_2)$ and any rate pair (R_1, R_2) on the boundary of $\mathcal{R}(D_1^*, D_2^*)$, with $R_1 \in \overline{\mathcal{R}}_1(D_1, D_2)$, we have

$$R_1 + R_2 \leq R_1^* + R_2^* + c_1(D_1, D_2) + c_2(D_1, D_2) \quad (53)$$

Moreover, for any $R_1^* \in \overline{\mathcal{R}}_1(D_1, D_2)$, there exists a rate pair $(R_1, R_2) \in \mathcal{R}(D_1^*, D_2^*)$ such that:

$$\begin{aligned} R_1 &= R_1^* \\ R_2 &\leq R_2^* + c_1(D_1, D_2) + c_2(D_1, D_2) \end{aligned} \quad (54)$$

Theorem 4. For any rate pair (R_1, R_2) on the boundary of $\mathcal{R}(D_1^*, D_2^*)$, with $R_1 \in \overline{\mathcal{R}}_1(D_1, D_2)$, there exists a rate pair $(R_1, R_2) \in \mathcal{R}(D_1^*, D_2^*)$ such that:

$$\begin{aligned} R_1 &\leq R_1^* + c_1(D_1, D_2) \\ R_2 &\leq R_2^* + c_2(D_1, D_2) \end{aligned} \quad (55)$$

Acknowledgment

The authors are grateful to Prof. Rami Zamir for useful discussions.

Appendix A - Universal Slepian-Wolf Coding

In this appendix we describe the universal Slepian-Wolf decoder used in our coding scheme. The following results are similar to those of [17]. For convenience, we omit the notation of the conditioning on the dither variables Z_1 and Z_2 . The results below can be applied for any realization of these continuous variables. Remember that our coding scheme, unlike the scheme presented in [6], requires only one realization of Z_1 and Z_2 in each round.

We consider the Slepian-Wolf setting for two correlated memoryless sources $(Y_1, Y_2) \sim P_{Y_1, Y_2}$. We assume that $Y_1 \in \mathcal{Y}_1$ and $Y_2 \in \mathcal{Y}_2$, where \mathcal{Y}_1 and \mathcal{Y}_2 are finite alphabets. A $(2^{nR_1}, 2^{nR_2}, n)$ source code is a block code of length n consisting of two encoders f_1, f_2 ,

$$f_1 : \mathcal{Y}_1^n \rightarrow I_{M_1}, \quad f_2 : \mathcal{Y}_2^n \rightarrow I_{M_2} \quad (\text{A.1})$$

and a decoder g

$$g : I_{M_1} \times I_{M_2} \rightarrow \mathcal{Y}_1^n \times \mathcal{Y}_2^n. \quad (\text{A.2})$$

where $M_j = 2^{nR_j}$, $j = 1, 2$. The probability of error of the code is defined as

$$P_e(n) \triangleq \Pr\{g(f_1(\mathbf{Y}_1), f_2(\mathbf{Y}_2)) \neq (\mathbf{Y}_1, \mathbf{Y}_2)\} \quad (\text{A.3})$$

We will prove the following result:

Theorem 5. *Let (R_1, R_2) be given. Then, there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$ Slepian-Wolf source codes with probability of error $P_e(n) \rightarrow 0$ as $n \rightarrow \infty$ for every memoryless source that satisfies Eq. (1).*

Proof. Throughout the proof, the cardinality of a set \mathcal{A} is denoted by $|\mathcal{A}|$. The empirical joint entropy $H_{\mathbf{y}_1, \mathbf{y}_2}(Y_1, Y_2)$ and the empirical conditional entropy $H_{\mathbf{y}_1, \mathbf{y}_2}(Y_1|Y_2)$ induced by the sequences $\mathbf{y}_1 \in \mathcal{Y}_1^n$, $\mathbf{y}_2 \in \mathcal{Y}_2^n$ are defined as

$$H_{\mathbf{y}_1, \mathbf{y}_2}(Y_1, Y_2) \triangleq - \sum_{y_1 \in \mathcal{Y}_1} \sum_{y_2 \in \mathcal{Y}_2} P_{\mathbf{y}_1, \mathbf{y}_2}(y_1, y_2) \log P_{\mathbf{y}_1, \mathbf{y}_2}(y_1, y_2) \quad (\text{A.4})$$

$$H_{\mathbf{y}_1, \mathbf{y}_2}(Y_1|Y_2) \triangleq - \sum_{y_1 \in \mathcal{Y}_1} \sum_{y_2 \in \mathcal{Y}_2} P_{\mathbf{y}_1, \mathbf{y}_2}(y_1, y_2) \log P_{\mathbf{y}_1, \mathbf{y}_2}(y_1|y_2) \quad (\text{A.5})$$

where $P_{\mathbf{y}_1, \mathbf{y}_2}(y_1, y_2)$, $P_{\mathbf{y}_1, \mathbf{y}_2}(y_1|y_2)$ are the empirical joint and conditional distribution functions, respectively, induced by \mathbf{y}_1 and \mathbf{y}_2 (see [18, Chap. 11]).

To prove the theorem, we use the following random-binning mechanism:

- *Codebook generation:* Assign every $\mathbf{y}_1 \in \mathcal{Y}_1^n$ to one of 2^{nR_1} bins independently according to a uniform distribution on $\{1, 2, \dots, 2^{nR_1}\}$. Similarly, randomly assign every $\mathbf{y}_2 \in \mathcal{Y}_2^n$ to one of 2^{nR_2} bins. Reveal the assignments f_1 and f_2 to the encoders and the decoder.
- *Encoding:* User j sends the index of the bin to which \mathbf{Y}_j belongs, $j = 1, 2$.
- *Decoding:* Given the received index pair $(T_1 = f_1(\mathbf{Y}_1), T_2 = f_2(\mathbf{Y}_2))$, the decoder uses the Minimum Joint Entropy (MJE) decoder: Choose the pair $(\mathbf{y}'_1, \mathbf{y}'_2) : f_1(\mathbf{y}'_1) = T_1, f_2(\mathbf{y}'_2) = T_2$ which minimizes the empirical joint entropy induced by $(\mathbf{y}'_1, \mathbf{y}'_2)$, $H_{\mathbf{y}'_1, \mathbf{y}'_2}(Y_1, Y_2)$.

Define the following events:

$$\begin{aligned}
E_0 &= \{(\mathbf{Y}_1, \mathbf{Y}_2) \notin A_\epsilon^n\} \\
E_1 &= \{(\mathbf{Y}_1, \mathbf{Y}_2) \in A_\epsilon^n\} \cap \{\exists \mathbf{y}'_1 \neq \mathbf{Y}_1 : f_1(\mathbf{y}'_1) = T_1 \text{ and } H_{\mathbf{y}'_1, \mathbf{Y}_2}(Y_1, Y_2) \leq H_{\mathbf{Y}_1, \mathbf{Y}_2}(Y_1, Y_2)\} \\
E_2 &= \{(\mathbf{Y}_1, \mathbf{Y}_2) \in A_\epsilon^n\} \cap \{\exists \mathbf{y}'_2 \neq \mathbf{Y}_2 : f_2(\mathbf{y}'_2) = T_2 \text{ and } H_{\mathbf{Y}_1, \mathbf{y}'_2}(Y_1, Y_2) \leq H_{\mathbf{Y}_1, \mathbf{Y}_2}(Y_1, Y_2)\} \\
E_{12} &= \{(\mathbf{Y}_1, \mathbf{Y}_2) \in A_\epsilon^n\} \cap \{\exists (\mathbf{y}'_1, \mathbf{y}'_2) : \mathbf{y}'_1 \neq \mathbf{Y}_1, \mathbf{y}'_2 \neq \mathbf{Y}_2, f_1(\mathbf{y}'_1) = T_1, f_2(\mathbf{y}'_2) = T_2 \\
&\quad \text{and } H_{\mathbf{y}'_1, \mathbf{y}'_2}(Y_1, Y_2) \leq H_{\mathbf{Y}_1, \mathbf{Y}_2}(Y_1, Y_2)\}
\end{aligned} \tag{A.6}$$

where A_ϵ^n , $\epsilon > 0$, is the strongly typical set with respect to the source P_{Y_1, Y_2} (see [18, Eq. 10.107]). Remember that \mathbf{Y}_1 , \mathbf{Y}_2 , f_1 and f_2 are random. Obviously,

$$H_{\mathbf{y}'_1, \mathbf{y}'_2}(Y_1, Y_2) \leq H_{\mathbf{y}_1, \mathbf{y}_2}(Y_1, Y_2) \Leftrightarrow H_{\mathbf{y}'_1, \mathbf{y}'_2}(Y_1|Y_2) \leq H_{\mathbf{y}_1, \mathbf{y}_2}(Y_1|Y_2) \tag{A.7}$$

where $H_{\mathbf{y}'_1, \mathbf{y}'_2}(Y_1|Y_2)$, $H_{\mathbf{y}_1, \mathbf{y}_2}(Y_1|Y_2)$ are the empirical conditional entropies induced by $(\mathbf{y}'_1, \mathbf{y}'_2)$ and $(\mathbf{y}_1, \mathbf{y}_2)$, respectively. We have an error if there is another pair of sequences in the same bin such that the empirical joint entropy induced by this pair is smaller than the empirical joint entropy induced by $(\mathbf{Y}_1, \mathbf{Y}_2)$. Hence,

$$\begin{aligned}
\bar{P}_e(n) &\leq \Pr\{E_0 \cup E_1 \cup E_2 \cup E_{12}\} \\
&\leq \Pr\{E_0\} + \Pr\{E_1\} + \Pr\{E_2\} + \Pr\{E_{12}\}
\end{aligned} \tag{A.8}$$

where $\bar{P}_e(n) \triangleq \mathbb{E}[P_e(n)]$ is the expected probability of error where the expectation is taken with respect to the random choice of the code. The first inequality follows from the fact that we treat E_0 as error event and the second inequality is due to the union bound. We

first consider E_0 . By the asymptotic equipartition property (AEP), $\Pr\{E_0\} \rightarrow 0$ and hence for n sufficiently large, $\Pr\{E_0\} < \epsilon$. To bound $\Pr(E_1)$, we have

$$\begin{aligned}
\Pr(E_1) &= \sum_{(\mathbf{y}_1, \mathbf{y}_2) \in A_\epsilon^n} P(\mathbf{y}_1, \mathbf{y}_2) \\
&\quad \cdot \Pr\{\exists \mathbf{y}'_1 \neq \mathbf{y}_1 : f_1(\mathbf{y}'_1) = f_1(\mathbf{y}_1) \text{ and } H_{\mathbf{y}'_1, \mathbf{y}_2}(Y_1|Y_2) \leq H_{\mathbf{y}_1, \mathbf{y}_2}(Y_1|Y_2)\} \\
&\leq \sum_{(\mathbf{y}_1, \mathbf{y}_2) \in A_\epsilon^n} P(\mathbf{y}_1, \mathbf{y}_2) \sum_{\mathbf{y}'_1 \in B(\mathbf{y}_1, \mathbf{y}_2)} \Pr\{f_1(\mathbf{y}'_1) = f_1(\mathbf{y}_1)\} \\
&= \sum_{(\mathbf{y}_1, \mathbf{y}_2) \in A_\epsilon^n} P(\mathbf{y}_1, \mathbf{y}_2) 2^{-nR_1} |B(\mathbf{y}_1, \mathbf{y}_2)| \tag{A.9}
\end{aligned}$$

where the set $B(\mathbf{y}_1, \mathbf{y}_2)$ is defined as

$$B(\mathbf{y}_1, \mathbf{y}_2) \triangleq \{\mathbf{y}'_1 : H_{\mathbf{y}'_1, \mathbf{y}_2}(Y_1|Y_2) \leq H_{\mathbf{y}_1, \mathbf{y}_2}(Y_1|Y_2)\} \tag{A.10}$$

and the last equality simply follows from the definition of the random-binning coding scheme. Using the method of types (see [18, Chap. 10-11]), we have

$$\begin{aligned}
|B(\mathbf{y}_1, \mathbf{y}_2)| &= \sum_{\mathbf{y}'_1 \in B(\mathbf{y}_1, \mathbf{y}_2)} 1 \\
&= \sum_{V_{\mathbf{y}'_1|\mathbf{y}_2} \subseteq B(\mathbf{y}_1, \mathbf{y}_2)} |V_{\mathbf{y}'_1|\mathbf{y}_2}| \\
&\leq \sum_{V_{\mathbf{y}'_1|\mathbf{y}_2} \subseteq B(\mathbf{y}_1, \mathbf{y}_2)} 2^{n(H_{\mathbf{y}'_1, \mathbf{y}_2}(Y_1|Y_2) + \epsilon)} \\
&\leq \sum_{V_{\mathbf{y}'_1|\mathbf{y}_2} \subseteq B(\mathbf{y}_1, \mathbf{y}_2)} 2^{n(H_{\mathbf{y}_1, \mathbf{y}_2}(Y_1|Y_2) + \epsilon)} \\
&\leq (n+1)^{|\mathcal{Y}_1||\mathcal{Y}_2|} 2^{n(H_{\mathbf{y}_1, \mathbf{y}_2}(Y_1|Y_2) + \epsilon)} \\
&\leq (n+1)^{|\mathcal{Y}_1||\mathcal{Y}_2|} 2^{n(H(Y_1|Y_2) + 2\epsilon)} \tag{A.11}
\end{aligned}$$

where $V_{\mathbf{y}'_1|\mathbf{y}_2}$ is the conditional type of \mathbf{y}'_1 given \mathbf{y}_2 (see [18, Chap. 10]). The second equality follows from the fact that the event $\mathbf{y}'_1 \in B(\mathbf{y}_1, \mathbf{y}_2)$ depends only on the type $V_{\mathbf{y}'_1|\mathbf{y}_2}$. In the first inequality, we used the known upper bound on the size of the conditional type. The second inequality stems from the definition of $B(\mathbf{y}_1, \mathbf{y}_2)$. In the third inequality we used a known upper bound on the number of conditional types. The last inequality follows since (see [18, Chap. 10])

$$(\mathbf{y}_1, \mathbf{y}_2) \in A_\epsilon^n \Rightarrow H_{\mathbf{y}_1, \mathbf{y}_2}(Y_1|Y_2) \leq H(Y_1|Y_2) + \epsilon. \tag{A.12}$$

Therefore, we have

$$\begin{aligned} \Pr(E_1) &\leq \sum_{(\mathbf{y}_1, \mathbf{y}_2) \in A_\epsilon^n} P(\mathbf{y}_1, \mathbf{y}_2) 2^{-nR_1} |B(\mathbf{y}_1, \mathbf{y}_2)| \\ &\leq (n+1)^{|\mathcal{Y}_1||\mathcal{Y}_2|} 2^{-nR_1} 2^{n(H(Y_1|Y_2)+2\epsilon)} \end{aligned} \quad (\text{A.13})$$

where in the second inequality we used Eq. (A.11). Similarly, it can be shown that

$$\Pr(E_2) \leq (n+1)^{|\mathcal{Y}_1||\mathcal{Y}_2|} 2^{-nR_2} \cdot 2^{n(H(Y_2|Y_1)+2\epsilon)} \quad (\text{A.14})$$

and

$$\Pr(E_{12}) \leq (n+1)^{|\mathcal{Y}_1||\mathcal{Y}_2|} 2^{-n(R_1+R_2)} \cdot 2^{n(H(Y_1, Y_2)+2\epsilon)} \quad (\text{A.15})$$

Hence, taking $R_1 > H(Y_1|Y_2) + 2\epsilon$, $R_2 > H(Y_2|Y_1) + 2\epsilon$ and $R_1 + R_2 > H(Y_1, Y_2) + 2\epsilon$, we have $P(E_1) < \epsilon$, $P(E_2) < \epsilon$ and $P(E_{12}) < \epsilon$ for sufficiently large n . Since $\bar{P}_e(n) \leq 4\epsilon$, there exists at least one universal code (f_1^*, f_2^*, g^*) with $P_e(n) \leq 4\epsilon$. Thus, we can construct a sequence of universal codes with $P_e(n) \rightarrow 0$, and the proof of achievability is complete. \square

Remark. It can be shown that the universal decoder presented in the proof above also achieves the optimal error exponent.

Appendix B - Proof of Lemma 2

We now prove Lemma 2. We first show that the random vector $(Y_1 - Z_1, Y_2 - Z_2)$ is equivalent to the random vector $(X_1 + N_1, X_2 + N_2)$ where N_1, N_2 are independent of X_1, X_2 and of each other and $N_i \sim \mathcal{U}[-\sqrt{3D_i}, \sqrt{3D_i}]$, $i \in \{1, 2\}$. Therefore, the dithered quantization process can be viewed as passing X_1, X_2 through independent noisy memoryless channels $\hat{X}_1 = X_1 + N_1$ and $\hat{X}_2 = X_2 + N_2$, respectively. We start with the following conditional probability distribution.

$$f_{N_1, N_2|X_1, X_2}(N_1, N_2|X_1, X_2) = f(N_1|X_1) f(N_2|X_2) \quad (\text{B.1})$$

where we have defined $N_1 \triangleq Y_1 - Z_1 - X_1$, $N_2 \triangleq Y_2 - Z_2 - X_2$. The equality stems from the fact that $(Y_1 - Z_1 - X_1)$ is independent of X_2 given X_1 and $(Y_2 - Z_2 - X_2)$ is independent of X_1 given X_2 , since (Z_1, Z_2) are independent of (X_1, X_2) . In addition, it can be easily seen that for every value of X_i , N_i is uniformly distributed over $[-\sqrt{3D_i}, \sqrt{3D_i}]$. Therefore, N_i is independent of X_i and we have

$$f_{N_1, N_2|X_1, X_2}(N_1, N_2|X_1, X_2) = f(N_1) f(N_2) \quad (\text{B.2})$$

Lemma 2 follows directly from this result:

$$\begin{aligned}
\mathbb{E}[Y_i - Z_i] &= \mathbb{E}[X_i + N_i] &= \mathbb{E}[X_i] \\
\mathbb{E}[(Y_i - Z_i)^2] &= \mathbb{E}[(X_i + N_i)^2] &= \mathbb{E}[X_i^2] + D_i \\
\mathbb{E}[X_i(Y_i - Z_i)] &= \mathbb{E}[X_i(X_i + N_i)] &= \mathbb{E}[X_i^2] \\
\mathbb{E}[(Y_1 - Z_1)(Y_2 - Z_2)] &= \mathbb{E}[(X_1 + N_1)(X_2 + N_2)] &= \mathbb{E}[X_1 X_2] \\
\mathbb{E}[X_1(Y_2 - Z_2)] &= \mathbb{E}[X_1(X_2 + N_2)] &= \mathbb{E}[X_1 X_2] \\
\mathbb{E}[X_2(Y_1 - Z_1)] &= \mathbb{E}[X_2(X_1 + N_1)] &= \mathbb{E}[X_1 X_2]
\end{aligned}$$

Appendix C - Calculation of the Estimation Error

In this appendix we calculate the estimation error given in Eq. (49). The optimal linear estimator of X_1 given the vector \underline{Y} is:

$$\hat{X}_1 = \underline{Y} \cdot \frac{1}{|\Lambda|} \begin{pmatrix} |\Lambda| - D_1(\mathbb{E}[X_2^2] + D_2) \\ \mathbb{E}[X_1 X_2] D_1 \end{pmatrix} \quad (\text{C.1})$$

where

$$|\Lambda| = (\mathbb{E}[X_1^2] + D_1)(\mathbb{E}[X_2^2] + D_2) - \mathbb{E}[X_1 X_2]^2 \quad (\text{C.2})$$

The error of the optimal linear estimator is given by:

$$D_1^* = \mathbb{E}[X_1^2] - \mathbb{E}[\hat{X}_1^2] \quad (\text{C.3})$$

Calculating the second term:

$$\begin{aligned}
|\Lambda|^2 \mathbb{E} [\hat{X}_1^2] &= (|\Lambda| - D_1 (\mathbb{E}[X_2^2] + D_2))^2 \mathbb{E} [(Y_1 - Z_1)^2] \\
&\quad + \mathbb{E}[X_1 X_2]^2 D_1^2 \mathbb{E} [(Y_2 - Z_2)^2] \\
&\quad + 2 (|\Lambda| - D_1 (\mathbb{E}[X_2^2] + D_2)) \mathbb{E}[X_1 X_2] D_1 \mathbb{E} [(Y_1 - Z_1) (Y_2 - Z_2)] \\
&= (|\Lambda| - D_1 (\mathbb{E}[X_2^2] + D_2))^2 (\mathbb{E}[X_1^2] + D_1) \\
&\quad + \mathbb{E}[X_1 X_2]^2 D_1^2 (\mathbb{E}[X_2^2] + D_2) \\
&\quad + 2 (|\Lambda| - D_1 (\mathbb{E}[X_2^2] + D_2)) \mathbb{E}[X_1 X_2]^2 D_1 \\
&= (|\Lambda| - D_1 (\mathbb{E}[X_2^2] + D_2))^2 (\mathbb{E}[X_1^2] + D_1) \\
&\quad + \mathbb{E}[X_1 X_2]^2 D_1 (D_1 (\mathbb{E}[X_2^2] + D_2) + 2 (|\Lambda| - D_1 (\mathbb{E}[X_2^2] + D_2))) \\
&= (|\Lambda| - D_1 (\mathbb{E}[X_2^2] + D_2))^2 (\mathbb{E}[X_1^2] + D_1) \\
&\quad + \mathbb{E}[X_1 X_2]^2 D_1 (2|\Lambda| - D_1 (\mathbb{E}[X_2^2] + D_2)) \\
&= (|\Lambda| - D_1 (\mathbb{E}[X_2^2] + D_2)) ((|\Lambda| - D_1 (\mathbb{E}[X_2^2] + D_2)) (\mathbb{E}[X_1^2] + D_1) + \mathbb{E}[X_1 X_2]^2 D_1) \\
&\quad + |\Lambda| \mathbb{E}[X_1 X_2]^2 D_1 \\
&= (|\Lambda| - D_1 (\mathbb{E}[X_2^2] + D_2)) (|\Lambda| (\mathbb{E}[X_1^2] + D_1) - D_1 |\Lambda|) \\
&\quad + |\Lambda| \mathbb{E}[X_1 X_2]^2 D_1 \\
&= (|\Lambda| - D_1 (\mathbb{E}[X_2^2] + D_2)) |\Lambda| \mathbb{E}[X_1^2] \\
&\quad + |\Lambda| \mathbb{E}[X_1 X_2]^2 D_1 \\
&= |\Lambda|^2 \mathbb{E}[X_1^2] + |\Lambda| D_1 (\mathbb{E}[X_1 X_2]^2 - \mathbb{E}[X_1^2] (\mathbb{E}[X_2^2] + D_2)) \\
&= |\Lambda|^2 \mathbb{E}[X_1^2] - |\Lambda| D_1 (\mathbb{E}[X_1^2] (\mathbb{E}[X_2^2] + D_2) - \mathbb{E}[X_1 X_2]^2) \\
&= |\Lambda|^2 \mathbb{E}[X_1^2] - |\Lambda| D_1 (|\Lambda| - D_1 (\mathbb{E}[X_2^2] + D_2)) \tag{C.4}
\end{aligned}$$

where in the second equality we used the results of Lemma 2. Therefore, we have

$$\begin{aligned}
D_1^* &= \frac{|\Lambda| D_1 (|\Lambda| - D_1 (\mathbb{E}[X_2^2] + D_2))}{|\Lambda|^2} \\
&= D_1 \left(1 - \frac{D_1 (\mathbb{E}[X_2^2] + D_2)}{|\Lambda|} \right) \\
&= D_1 \left(1 - \frac{D_1 (\mathbb{E}[X_2^2] + D_2)}{(\mathbb{E}[X_1^2] + D_1) (\mathbb{E}[X_2^2] + D_2) - \mathbb{E}[X_1 X_2]^2} \right) \\
&= D_1 \frac{\mathbb{E}[X_1^2] (\mathbb{E}[X_2^2] + D_2) - \mathbb{E}[X_1 X_2]^2}{(\mathbb{E}[X_1^2] + D_1) (\mathbb{E}[X_2^2] + D_2) - \mathbb{E}[X_1 X_2]^2} \tag{C.5}
\end{aligned}$$

References

- [1] D. Slepian and J. K. Wolf, “Noiseless coding of correlated information sources,” *IEEE Trans. Inform. Theory*, vol. 19, pp. 471–480, Jul. 1973.
- [2] A. D. Wyner and J. Ziv, “The rate-distortion function for source coding with side information at the decoder,” *IEEE Trans. Inform. Theory*, vol. 22, pp. 1–10, Jan. 1976.
- [3] A. D. Wyner, “The rate-distortion function for source coding with side information at the decoder-II: General sources,” *Inform. Contr.*, vol. 38, pp. 60–80, 1978.
- [4] R. Ahlswede and J. Körner, “Source coding with side information and a converse for degraded broadcast channels,” *IEEE Trans. Inform. Theory*, vol. 21, no. 6, pp. 629–637, Nov. 1975.
- [5] T. Berger and R. W. Yeung, “Multiterminal source encoding with one distortion criterion,” *IEEE Trans. Inform. Theory*, vol. 35, no. 2, pp. 228–236, Mar. 1989.
- [6] R. Zamir and T. Berger, “Multiterminal source coding with high resolution,” *IEEE Trans. Inf. Theory*, vol. 45, pp. 106–117, Jan. 1999.
- [7] A. Wagner and V. Anantharam, “An improved outer bound for multiterminal source coding,” *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1919–1937, May 2008.
- [8] A. Wagner, S. Tavildar, and P. Viswanath, “Rate region of the quadratic gaussian two-encoder source-coding problem,” *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1938–1961, May 2008.
- [9] T. Berger and S. Y. Tung, “Encoding of correlated analog sources,” in *Proc. IEEE-USSR Joint Workshop on Information Theory*, pp. 7–10, 1975.
- [10] T. A. Courtade and T. Weissman, “Multiterminal source coding under logarithmic loss,” *IEEE Trans. Inform. Theory*, vol. 60, pp. 740–761, Jan. 2014.
- [11] Y. Kaspi and N. Merhav, “Zero-delay and causal single-user and multi-user lossy source coding with decoder side information,” *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 6931–6942, Nov. 2014.

- [12] J. Ziv, “On universal quantization,” *IEEE Trans. Inform. Theory*, vol. 31, pp. 344–347, May 1985.
- [13] R. Zamir and M. Feder, “On universal quantization by randomized uniform/lattice quantizer,” *IEEE Trans. Inform. Theory*, vol. 38, pp. 428–436, Mar. 1992.
- [14] R. Zamir and M. Feder, “Information rates of pre/post filtered dithered quantizers,” *IEEE Trans. Inform. Theory*, vol. 42, pp. 1340–1353, Sept. 1996.
- [15] J. Kieffer, “A unified approach to weak universal source coding,” *IEEE Trans. Inform. Theory*, vol. 24, pp. 674–682, Nov. 1978.
- [16] H. Gish and N. J. Pierce, “Asymptotically efficient quantization,” *IEEE Trans. Inform. Theory*, vol. 14, pp. 676–683, Sept. 1968.
- [17] I. Csiszar and J. Körner, “Towards a general theory of source networks,” *IEEE Trans. Inform. Theory*, vol. 26, pp. 155–165, Mar. 1980.
- [18] T. M. Cover and J. A. Thomas, “Elements of information theory,” *John Wiley & Sons*, 2nd Edition.